

VIRGINIE RONDEAU

DANIEL COMMENGES

Modélisation de la fragilité en survie

Journal de la société française de statistique, tome 143, n° 1-2 (2002),
p. 103-119

http://www.numdam.org/item?id=JSFS_2002__143_1-2_103_0

© Société française de statistique, 2002, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

MODÉLISATION DE LA FRAGILITÉ EN SURVIE

Virginie RONDEAU * et Daniel COMMENGES *

RÉSUMÉ

Cet article présente une revue des modèles à fragilités ou « frailty models ». L'accent est principalement mis sur les modèles à risque multiplicatif où une hypothèse de distribution gamma est faite pour la fragilité. Le modèle de base est un modèle à fragilités partagées par les sujets d'un même groupe. Un modèle plus général est un modèle à fragilités corrélées. Différentes approches ont été proposées pour ajuster ces modèles, en particulier l'utilisation de l'algorithme EM et celle de la vraisemblance pénalisée.

ABSTRACT

This article presents a review of frailty models. We particularly examine models with multiplicative hazard and with the assumption of gamma distribution for the frailty variable. The baseline model is a model with frailties shared by subjects from the same group. A more general model is a correlated frailty model. Different approaches have been proposed to fit these models, especially the use of the EM algorithm and the penalized likelihood.

1. Introduction

Les effets aléatoires, déjà utilisés dans les modèles d'analyse de composantes de la variance, ont fait leur apparition dans les modèles linéaires généraux à la fin des années 70 (Harville, 1977). Leur irruption dans les modèles de survie est presque contemporaine (Clayton, 1978). Ces modèles ont été appelés « frailty models » ce qui peut se traduire en bon (?) français par « modèles à fragilités ». Ce terme évoque dans le contexte de la fiabilité qu'une pièce peut être plus fragile qu'une autre et donc avoir un risque d'échec (« failure ») plus grand, dans le contexte épidémiologique qu'un sujet peut être plus fragile qu'un autre et donc avoir un risque de mort ou d'autre événement pathologique plus grand. L'impact de l'hétérogénéité de la population sur l'interprétation de la forme des fonctions de risques est important (Aalen, 1994). Il faut reconnaître que les modèles à fragilités ont eu moins d'impact sur les pratiques épidémiologiques que les modèles linéaires mixtes. Les raisons en sont diverses. L'une est que les effets aléatoires sont nécessaires dans les modèles linéaires mixtes pour traiter des données longitudinales tandis que les modèles de survie sont déjà des

* INSERM U330, 146 rue Léo Saigat, 33076 Bordeaux ; E-mail : Virginie.Rondeau@isped.u-bordeaux2.fr

modèles pour données longitudinales (événementielles). Une autre raison tient à la complexité plus grande de ces modèles par rapport aux modèles linéaires normaux. Cette complexité entraîne des difficultés théoriques pour démontrer les propriétés des estimateurs et d'autre part une plus grande difficulté pour trouver des algorithmes numériques efficaces. Les modèles à fragilités subissent la concurrence des approches marginales (Wei, Lin et Weissfeld, 1989) qui contournent ces difficultés. Cependant la modélisation de la fragilité est en elle-même intéressante et permet des interprétations que l'approche marginale ne permet pas. En raison de ces difficultés, des travaux ont été développés pour tester s'il existait réellement une hétérogénéité (Clayton et Cuzick, 1985 ; Commenges, 1995). Nous nous focaliserons ici sur l'objectif plus ambitieux d'ajuster un modèle à fragilités en faisant état de développements récents qui n'ont pas été présentés dans l'article de Hougaard (1995).

L'article sera organisé de la manière suivante. Dans un premier temps nous décrirons différentes situations de données hétérogènes. La section 3 décrit le modèle à fragilités partagées et les différentes méthodes d'estimation utilisées. Le modèle à fragilités corrélées est ensuite exposé dans la section 4. Nous concluons en évaluant l'impact de cette approche et les futures voies de recherche.

2. Domaines d'application et impact

Les modèles à fragilités peuvent être utilisés dans des applications assez différentes. Il y a une première distinction à faire entre les modèles multivariés et non-multivariés. Dans les premiers, on modélise de manière conjointe deux types d'événements différents ; par exemple on peut s'intéresser chez des sujets infectés par le VIH à la survenue de deux maladies opportunistes : le syndrome de Kaposi et la pneumocystose. À l'opposé, plusieurs domaines d'application des modèles à fragilités ne concernent qu'un type d'événement, et nous ne dirons pas qu'il s'agit de modèles multivariés même si les données sont corrélées (certains auteurs utilisent le vocable « multivarié » dans une acception moins spécifique). Parmi les situations comportant un seul type d'événement, les modèles à fragilités sont utilisés pour exprimer le fait que par rapport à un modèle de référence, les individus ont une variabilité. Il est souvent vrai qu'une part de cette variabilité peut être expliquée en introduisant des variables « explicatives » dans le modèle, tandis qu'une part ne peut l'être.

L'hétérogénéité pour un modèle de référence n'est pas nécessairement associée à une corrélation des données. Nous allons d'abord considérer un exemple d'hétérogénéité qui ne produit pas une telle corrélation. Nous aborderons ensuite le cas sans doute plus connu où il y a une hétérogénéité entre groupes provoquant une corrélation des données.

Prenons un exemple très simple d'un modèle de fertilité. Le modèle de base est d'attribuer une probabilité $\pi(t)$ de conception pour un couple à chaque cycle t ($t = 1, \dots, k$) où le couple essaie de concevoir ; il s'agit d'un modèle de survie en temps discret : le résultat est aléatoire (ou variable) mais la probabilité est la même pour tous les couples et reste constante dans le temps ($\pi(t) = \pi$).

Il semble naturel de penser que ce modèle est trop simple et en particulier l'hypothèse que la probabilité est la même pour tous les couples n'est pas tenable. Il est possible que l'on ait une variable explicative qui renseigne sur un antécédent d'infection des trompes chez la femme : cela permettrait de proposer un modèle donnant une probabilité π_1 aux femmes sans antécédent, π_2 à celles avec antécédent : ce modèle exprime une hétérogénéité entre femmes qui peut s'expliquer. Cependant, même pour les femmes n'ayant pas d'antécédent d'infection, la probabilité reste variable. Cette variabilité non expliquée peut être représentée par une variable aléatoire que l'on appellera « fragilité » (frailty); par exemple la probabilité pour le couple i s'écrira $\pi_i = Z_i \pi$ ou Z_i est une variable aléatoire d'espérance 1 et π est la probabilité de base (qui peut éventuellement dépendre de variables explicatives). On peut donc définir la notion d'hétérogénéité (et utiliser un modèle à fragilités) sans observer d'événements répétés ou groupés, donc sans données corrélées.

Il est intéressant d'examiner l'effet de l'hétérogénéité sur les probabilités marginales, c'est-à-dire celles que l'on peut estimer directement dans la population. Tietze (1968) a présenté des données de temps jusqu'à la conception pour des couples ayant arrêté la contraception : les taux de conception décroissent régulièrement et cela donne l'impression que la probabilité de concevoir décroît avec le temps. Aalen (1988) a analysé ces données avec un modèle à fragilités (un modèle beta-binomial) et montré qu'elles sont compatibles avec l'hypothèse de probabilité constante pour chaque couple, mais différentes entre couples. La décroissance des taux dans la population est due à un phénomène de sélection : les couples les plus fertiles conçoivent rapidement, ce qui fait qu'au bout d'un certain temps, il ne reste que les couples les moins fertiles (cf Figure 1).

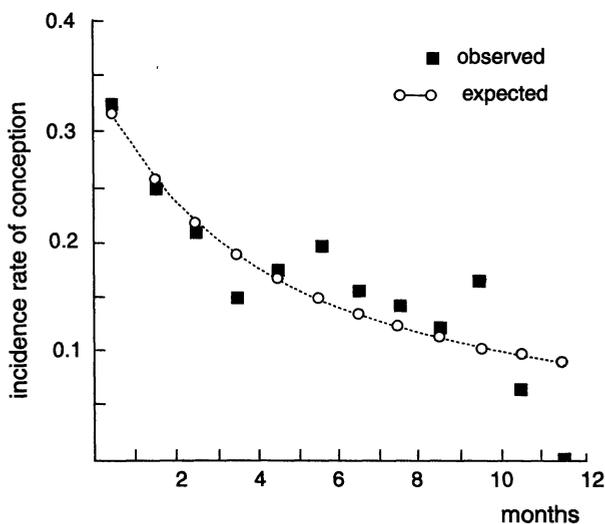


FIG. 1. – Taux d'incidence de conception par mois, observés et attendus dans un modèle à fragilités à probabilité constante, à partir des données de Tietze (1968). Figure reproduite de Aalen (1988) avec la permission de John Wiley and Sons.

Nous nous intéresserons surtout dans la suite à des modèles de survie en temps continu. Continuons l'étude de l'hétérogénéité sans corrélation, nous modélisons la fonction de risque pour l'individu i par $\lambda_i(t) = Z_i \lambda_0(t)$, où $\lambda_0(t)$ est la fonction de risque de base, les Z_i sont des variables aléatoires indépendantes et de même loi, avec $E(Z_i) = 1$ et $\text{var}(Z_i) = \theta$ pour tout i ; θ mesure le degré d'hétérogénéité. On peut montrer que la fonction de risque marginal est $\mu(t) = E[Z_i | T_i \geq t] \lambda_0(t)$, c'est-à-dire le risque parmi les survivants ($T_i \geq t$). $E[Z_i | T_i \geq t]$ (qui ne dépend pas de i) est en général décroissant, ce qui rappelle l'exemple de la fertilité : si $\lambda_0(t)$ est constant, $\mu(t)$ sera décroissant.

L'hétérogénéité a aussi un effet sur l'évolution du risque relatif en fonction du temps. En particulier, si on part d'un modèle de référence à risques proportionnels, c'est-à-dire à risque relatif constant avec le temps ($r = \exp(\beta)$ pour une variable X_i binaire), le modèle à fragilités s'écrit $\lambda_i(t) = Z_i \lambda_0(t) \exp(\beta X_i)$, le risque relatif apparent décroît en fonction du temps en présence d'hétérogénéité. Cela est dû également à un phénomène de sélection : les plus fragiles du groupe à risque sont éliminés plus rapidement ; au bout d'un certain temps, le groupe le plus à risque est constitué d'individus moins fragiles que le groupe le moins à risque. Si la fragilité Z_i suit une distribution gamma, le risque relatif tend vers 1 ; pour d'autres distributions le risque relatif peut même s'inverser ! Si conditionnellement à la fragilité, le modèle est à risques proportionnels, il ne l'est plus marginalement. Ainsi que nous l'avons mentionné dans la section 2, le risque relatif marginal décroît en fonction du temps à partir de la vraie valeur. Un modèle à risques proportionnels donne comme estimation une moyenne des valeurs de ce risque, qui est donc comprise entre la vraie valeur et 1 (pour une valeur positive des coefficients de régression). Afin d'illustrer ce phénomène, considérons le cas où les risques de base dans deux groupes sont respectivement $\lambda_0(t)$ et $r \lambda_0(t)$, avec r le risque relatif constant pour un facteur de risque spécifique. Les risques relatifs individuels sont respectivement $Z_1 \lambda_0(t)$ et $Z_2 r \lambda_0(t)$, pour des variables de fragilité Z_1 et Z_2 . Si l'on considère des variables de fragilité qui suivent une loi gamma, de moyenne 1 et de variance θ , alors les fonctions de risque marginal sont respectivement égales à $\mu_1(t) = \frac{\lambda_0(t)}{1 + \theta \Lambda_0(t)}$ et $\mu_2(t) = \frac{r \lambda_0(t)}{1 + r \theta \Lambda_0(t)}$. On obtient alors l'expression suivante pour le risque relatif dans la population : $RR_{\text{marginal}} = \frac{r(1 + \theta \Lambda_0(t))}{(1 + r \theta \Lambda_0(t))}$. La figure 2 illustre la décroissance au cours du temps du risque relatif marginal ou risque relatif dans la population pour $\theta = 2$, $r = 4$ et une fonction de risque de base constante égale à 1. Notons que le temps en abscisse s'interprète ici également comme la fonction de risque cumulé.

Si l'on a un modèle de référence satisfaisant comme dans le cas de la fertilité, il est souhaitable d'utiliser un modèle à fragilités pour rendre compte de l'hétérogénéité non expliquée de la population. Cependant, si ces variables de fragilité sont indépendantes, il y a peu d'information pour estimer leur variance. Les cas les plus intéressants pour l'analyse sont ceux où l'on observe plusieurs événements partageant la même variable de fragilité. Cela peut

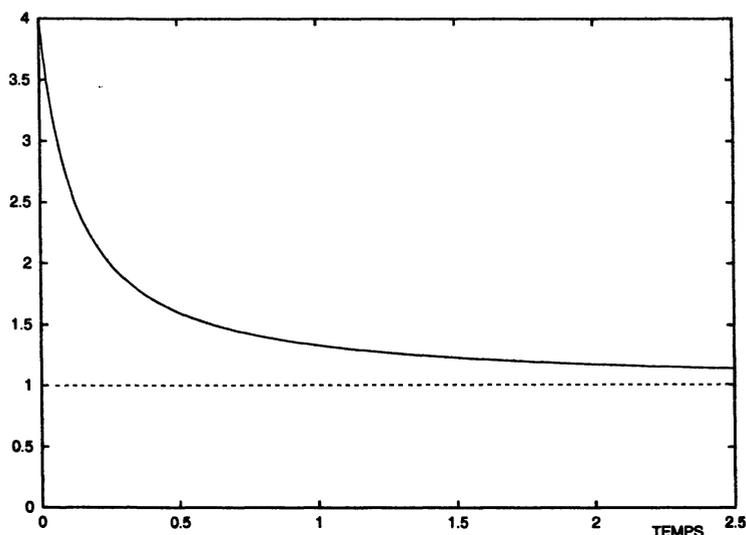


FIG. 2. – Risque relatif dans la population ou risque marginal, pour une vraie valeur du risque relatif = 4 et $\Lambda_0(t) = t$.

se produire en épidémiologie dans deux types de situations : événements répétés (exemple des cathéters développé dans la section 3.4) et données groupées. Dans l'exemple des cathéters, chaque sujet fournit deux observations de temps de «survie» d'un cathéter : les données sont répétées dans le temps. Pour les données groupées, on peut citer le cas de la cohorte Paquid, dans laquelle on s'intéresse en particulier à l'effet de différents facteurs de risque dans la survenue d'une démence. Les sujets ont été tirés au sort dans différentes communes de Dordogne et de Gironde (Rondeau *et al.*, 2002). Il est possible que les sujets de différentes communes ne suivent pas le même modèle. Une modélisation naturelle est de supposer que seuls les sujets d'une même commune suivent le même modèle. Dans un tel modèle d'homogénéité intra-groupe et d'hétérogénéité inter-groupe, il y a une corrélation marginale positive entre les durées de vie des sujets d'un même groupe. Les modèles à fragilités partagées sont adaptés à cette situation. Des structures de corrélation plus complexes, que l'on rencontre en particulier en génétique, peuvent être traitées par des modèles à fragilités corrélées.

3. Modèles à fragilités partagées et inférence

Nous considérons des modèles où la fonction de risque dépend en partie d'une variable aléatoire qui agit multiplicativement sur le risque, de sorte qu'une valeur élevée de cette variable augmente le risque.

Nous traitons ici uniquement le cas des données censurées à droite. Pour le $j^{\text{ième}}$ individu ($j = 1, \dots, n_i$) du $i^{\text{ième}}$ groupe ($i = 1, \dots, G$), on pose T_{ij} les

temps de survie et C_{ij} les temps de censure correspondants. Les observations sont $Y_{ij} = \min(T_{ij}, C_{ij})$ et les indicateurs d'événement $\delta_{ij} = I_{\{T_{ij} \leq C_{ij}\}}$.

Le modèle à fragilités partagées (ou « *shared frailty model* ») spécifie que la fonction de risque conditionnelle à la variable de fragilité est :

$$\lambda_{ij}(t|Z_i) = Z_i \lambda_0(t) \exp(\beta' X_{ij}) \quad (1)$$

où $\lambda_0(t)$ est la fonction de risque de base ; $X_{ij} = (X_{1ij}, \dots, X_{pij})'$ représente le vecteur des variables explicatives pour le $j^{\text{ième}}$ individu du groupe i , β est le vecteur correspondant des paramètres de régression, et les Z_i sont des variables aléatoires non observées (ou variables de fragilité) et partagées par tous les individus d'un même groupe. Dans le cas particulier où $n_i = 1$, le modèle (1) correspond à un modèle de surdispersion ou de fragilité pure qui a été discuté dans la section 2. Nous considérerons par la suite le cas de modèles à fragilités gamma, c'est-à-dire pour lesquels les effets aléatoires Z_i sont indépendants et identiquement distribués, de moyenne 1 et de variance inconnue θ et ont une densité gamma $f_Z(z) = \frac{z^{(1/\theta-1)} \exp\{-z/\theta\}}{\Gamma(1/\theta)\theta^{1/\theta}}$, la plus utilisée actuellement de part ses propriétés mathématiques. Le choix de la distribution gamma a été discuté par Hougaard (1995). D'autres distributions ont été étudiées, la distribution log-normale ou la distribution positive stable, proposée par Hougaard (1986).

3.1. L'algorithme EM

Plusieurs algorithmes ont été proposés pour estimer les différents paramètres d'intérêts qui sont : les coefficients β , la fonction de risque de base et la variance θ des effets aléatoires, ou pour prédire les effets aléatoires. Nous exposons ici l'algorithme proposé par Klein (1992). Plusieurs variantes de cet algorithme ont été proposées (Nielsen *et al.*, 1992 ; Parner, 1997).

Si les effets aléatoires Z_i sont observés, la log-vraisemblance complète ou conjointe pour un modèle à fragilités gamma et des données potentiellement censurées à droite s'écrit :

$$\begin{aligned} & \sum_{i=1}^G \sum_{j=1}^{n_i} \log \{ \lambda_{ij}(Y_{ij}|Z_i)^{\delta_{ij}} \times S_{ij}(Y_{ij}|Z_i) \times f(Z_i) \} \\ & = \sum_{i=1}^G \sum_{j=1}^{n_i} \{ \delta_{ij} (\log \lambda_0(Y_{ij}) + \log Z_i + \beta' X_{ij}) \\ & \quad - \Lambda_0(Y_{ij}) Z_i \exp(\beta' X_{ij}) + \log(f(Z_i)) \} \\ & = L_1(\theta) + L_2(\beta, \Lambda_o(\cdot)) \end{aligned} \quad (2)$$

avec

$$L_1(\theta) = -G \left\{ \frac{1}{\theta} \log \theta + \log \Gamma\left(\frac{1}{\theta}\right) \right\} + \sum_{i=1}^G \left\{ \left(\frac{1}{\theta} + m_i - 1 \right) \log Z_i - \frac{Z_i}{\theta} \right\} \quad (3)$$

$$L_2(\beta, \Lambda_o(\cdot)) = \sum_{i=1}^G \left\{ \sum_{j=1}^{n_i} \{ \delta_{ij} (\log \lambda_0(Y_{ij}) + \beta' X_{ij}) - \Lambda_0(Y_{ij}) Z_i \exp(\beta' X_{ij}) \} \right\} \quad (4)$$

où $\Lambda_0(\cdot)$ est la fonction de risque cumulé de base et $S_{ij}(\cdot | Z_i)$ est la fonction de survie.

On obtient, par intégration directe de la vraisemblance conjointe par rapport à Z , la log-vraisemblance marginale :

$$\begin{aligned} l(\lambda_0(\cdot), \beta, \theta) = \sum_{i=1}^G \left\{ \sum_{j=1}^{n_i} \delta_{ij} \{ \beta' X_{ij} + \ln(\lambda_0(Y_{ij})) \} \right. \\ \left. - (1/\theta + m_i) \ln \left[1 + \theta \sum_{j=1}^{n_i} \Lambda_0(Y_{ij}) \exp(\beta' X_{ij}) \right] \right. \\ \left. + I_{\{m_i \neq 0\}} \sum_{k=1}^{m_i} [\ln(1 + \theta(m_i - k))] \right\} \quad (5) \end{aligned}$$

avec $m_i = \sum_{j=1}^{n_i} I_{\{\delta_{ij}=1\}}$ le nombre d'événements observés dans le $i^{\text{ème}}$ groupe.

Donc, la vraisemblance prend une forme analytique simple (il n'y a pas d'intégrale à calculer numériquement) : notons que c'est une propriété particulière de la distribution gamma. Les effets aléatoires peuvent être vus comme des variables explicatives non observées ou données manquantes. Ceci a conduit à l'utilisation de l'algorithme EM comme procédure d'estimation. L'algorithme EM alterne entre deux étapes, une étape de maximisation (étape M) et l'autre d'espérance (étape E).

(Pour une valeur fixée de θ)

1. Des valeurs initiales sont attribuées à β et $\Lambda_0(\cdot)$, en fixant $Z_i = 1$.

2. Etape E :

Z_i est remplacé par son espérance conditionnelle aux données, pour des valeurs courantes de β et de λ_0 , c'est-à-dire

$$E(Z_i | \text{données}) = \frac{1/\theta + m_i}{1/\theta + \sum_{j=1}^{n_i} \hat{\Lambda}_0(t) \exp(\beta' X_{ij})}$$

3. Etape M :

Cette étape nécessite respectivement une maximisation de (3) et de (4) par rapport aux paramètres inconnus θ et β . Pour estimer β , on effectue une maximisation de la log-vraisemblance $L_2(\beta, \Lambda_o(\cdot))$. Cette log-vraisemblance

contient le paramètre de nuisance $\Lambda_0(\cdot)$; il va être remplacé par l'estimateur de Breslow, soit :

$$\hat{\Lambda}_0(t) = \sum_{T_{(i)} \leq t} \frac{d_{(i)}}{\sum_{k \in R(T_{(i)})} \hat{Z}_k \exp(\beta' X_k)} \quad (6)$$

avec $T_{(i)}$ le $i^{\text{ième}}$ temps d'événement observé, $R(T_{(i)})$ l'ensemble des indices des sujets à risque au temps $T_{(i)}$ et $d_{(i)}$ le nombre d'événements au temps $T_{(i)}$ et \hat{Z}_k la valeur estimée de la variable de fragilité pour l'individu k . En remplaçant (6) dans (4) et en s'appuyant sur les arguments de Johansen (1983), on obtient une log-vraisemblance profilée qui prend la forme d'une log-vraisemblance partielle de Cox classique, dans laquelle on ajoute une nouvelle covariable $\log \hat{Z}_i$ de coefficient connu égal à 1 :

$$\sum_{i=1}^D \left\{ S_{(i)} \beta - d_{(i)} \log \sum_{k \in R(T_{(i)})} \hat{Z}_k \exp(\beta' X_k) \right\} \quad (7)$$

où D est le nombre de temps de décès, et $S_{(i)}$ est la somme des covariables des individus décédés au temps $T_{(i)}$.

Ces étapes 2 et 3 sont répétées jusqu'à convergence.

Une fois la convergence obtenue, les estimateurs des variances des paramètres $\hat{\beta}$ et $\hat{\theta}$ sont obtenus par inversion de la matrice d'information de la log-vraisemblance marginale (5).

Nielsen *et al.* (1992) proposent un autre algorithme EM similaire mais plus rapide. Ils choisissent d'estimer θ en dehors des étapes E et M. Pour cela ils utilisent une maximisation de la log-vraisemblance profilée $l(\theta) = l(\hat{\lambda}_0(\theta), \hat{\beta}(\theta), \theta)$ obtenue à partir de l'équation (5). Néanmoins, ces algorithmes EM restent longs.

Une macro SAS, écrite par Klein (disponible sur le site <http://www.biostat.mcw.edu/>) utilise une estimation des paramètres par l'algorithme EM, dans un modèle à fragilités partagées. Cette macro traite des données éventuellement censurées à droite, mais ne traite pas le cas des données tronquées à gauche.

3.2. L'approche par vraisemblance pénalisée (sur les variables de fragilité)

Therneau et Grambsch (2001) proposent un algorithme différent pour estimer les paramètres dans un modèle à fragilités gamma partagées. Pour cela, il est préférable d'écrire le modèle (1) en remplaçant Z_i par $\exp(Z_i^*)$:

$$\lambda_{ij}(t|Z_i^*) = \lambda_0(t) \exp(\beta' X_{ij} + Z_i^*) \quad (8)$$

Ils proposent de maximiser une log-vraisemblance partielle pénalisée, à la fois sur β et sur Z^* , dans laquelle un terme de pénalisation $-g(Z^*; \theta)$ est introduit afin d'éviter de larges différences entre les variables de fragilité :

$$ppl = pl(\beta, Z^*; \text{données}) - g(Z^*; \theta) \quad (9)$$

avec, $pl(\beta, Z^*; \text{données})$ la log-vraisemblance partielle de Cox usuelle et $g(Z^*; \theta)$ la fonction de pénalisation qui restreint les valeurs de Z^* . Le paramètre θ , qui correspond à la variance des effets aléatoires Z_i , peut être vu comme un paramètre de lissage, pré-spécifié ou estimé.

Ils démontrent que la solution obtenue dans un modèle à fragilités gamma partagées par une vraisemblance partielle pénalisée (équation 9) ayant pour fonction de pénalisation $g(Z^*; \theta) = -1/\theta \sum_{i=1}^G [Z_i^* - \exp(Z_i^*)]$ coïncide avec la solution obtenue par l'algorithme EM. Le choix de la fonction de pénalisation est lié à la densité des effets aléatoires Z_i^* . En effet, on peut remarquer que le log de la densité de Z^* (avec $Z = e^{Z^*}$ qui suit une loi gamma) est égal à $(Z^* - \exp Z^*)/\theta$ plus des termes qui ne dépendent pas de Z^* . Ainsi, le terme de pénalisation $-g(Z^*; \theta)$ prendra un poids important lorsque l'on aura des valeurs extrêmes pour Z_i , et donc une faible valeur de la densité.

Puisque la fonction de pénalisation ne fait pas intervenir β , on a $\partial ppl/\partial \beta = \partial pl/\partial \beta$. Ainsi, la fonction de score pour β est la même que celle utilisée dans un modèle à risques proportionnels dans lequel Z^* est traitée comme une variable « offset ». Puis, on peut montrer que l'équation de score pour Z_i^* s'écrit : $\frac{\partial ppl}{\partial Z_i^*} = [m_i - \sum_{j=1}^{n_i} \hat{\Lambda}_0(t) Z_i^*] - \frac{\partial g(Z_i^*, \theta)}{\partial Z_i^*} = 0$.

La variance des $p + G$ paramètres (soient p coefficients de régression et G effets aléatoires) peut être estimée par $H^{-1}IH^{-1}$ ou directement par H^{-1} , où H représente la matrice hessienne de la log-vraisemblance pénalisée, i.e. $H = H(\beta, Z^*) = I + \begin{pmatrix} 0 & 0 \\ 0 & g'' \end{pmatrix}$ et $I = I(\beta, Z^*)$ est la matrice d'information usuelle pour un modèle à risques proportionnels (Gray, 1992).

Ensuite, la variance θ des effets aléatoires est estimée par maximisation d'une log-vraisemblance marginale profilée (comme dans l'algorithme EM), $l(\theta) = l(\hat{\lambda}_0(\theta), \hat{\beta}(\theta), \theta)$ obtenue à partir de l'équation (5). Si l'on connaît $ppl(\theta) = ppl(\hat{\beta}(\theta), \hat{Z}^*(\theta), \theta)$ la log-vraisemblance pénalisée partielle profilée, on peut en déduire facilement une valeur de $l(\theta)$ par la relation suivante :

$$l(\theta) = ppl(\theta) + \sum_{i=1}^G \left\{ 1/\theta - (1/\theta + m_i) + 1/\theta \log(1/\theta) + \log \left(\frac{\Gamma(1/\theta + m_i)}{\Gamma(1/\theta)} \right) \right\}$$

Therneau et Grambsch n'obtiennent pas directement d'estimateur de la variance de $\hat{\theta}$; ils proposent d'utiliser pour cela une méthode graphique qui consiste à tracer la log-vraisemblance profilée $l(\hat{\lambda}_0(\theta), \hat{\beta}(\theta), \theta)$ pour différentes valeurs de θ .

En résumé, la procédure générale d'estimation des paramètres utilise deux boucles imbriquées l'une dans l'autre. Tout d'abord, pour un θ fixé, l'algorithme de Newton-Raphson permet de maximiser la log-vraisemblance pénalisée pour $\hat{\beta}(\theta)$ et $\hat{Z}^*(\theta)$, et fournit une valeur correspondante pour $ppl(\theta)$. Puis la boucle extérieure maximise la log-vraisemblance profilée pour estimer θ . Cette procédure d'estimation par vraisemblance pénalisée est disponible sous S-plus et R, pour des modèles à fragilités partagées avec une densité gamma pour Z ou une densité gaussienne. Une approche similaire a été développée par Ripatti et Palmgren (2000) dans le cas de distributions multivariées normales pour les effets aléatoires.

Ha *et al.* (2001) ont proposé récemment d'utiliser une estimation par vraisemblance hiérarchique (Lee et Nelder, 1996) dans les modèles à fragilités partagées. Cette approche est en fait très proche de la méthode d'estimation par vraisemblance partielle pénalisée de Therneau et Grambsch (2001). Notons aussi qu'une approche bayésienne a été proposée par Clayton (1991).

3.3. L'approche par vraisemblance pénalisée (sur la fonction de risque)

Les méthodes d'estimation exposées précédemment ne fournissent pas un estimateur lisse de la fonction de risque, or il est raisonnable de contraindre la fonction de risque à être continue et à avoir de faibles variations locales. De plus à partir de l'estimation de la fonction de risque de base $\lambda_0(t)$ dans un modèle à fragilités sans variables explicatives, on peut en déduire la fonction de risque marginal $\mu(t) = E[Z_i | T_i \geq t] \lambda_0(t)$; elle représente dans le cas particulier où l'âge est choisi comme temps de base, l'incidence de la maladie en fonction de l'âge. Nous avons donc proposé une nouvelle approche semi-paramétrique pour estimer les paramètres β, θ et la fonction de risque $\lambda_0(t)$ (Rondeau *et al.*, 2002). Nous utilisons également une vraisemblance pénalisée mais cette fois et contrairement à l'approche de Therneau (2001), la pénalisation opère sur les fluctuations de la fonction de risque, supposée lisse et non pas sur les effets aléatoires. Ainsi, nous pénalisons la log-vraisemblance par un terme qui prend des valeurs élevées pour des fonctions de risque irrégulières (O'Sullivan, 1988; Joly *et al.*, 1998). On définit la log-vraisemblance pénalisée par :

$$pl(\lambda_0(\cdot), \beta, \theta) = l(\lambda_0(\cdot), \beta, \theta) - \kappa \int_0^\infty \lambda_0''^2(t) dt \quad (10)$$

où $l(\lambda_0(\cdot), \beta, \theta)$ est la log-vraisemblance marginale définie en (5). Cette expression représente un compromis entre un ajustement sur les données, représentée par $l(\cdot)$ et une contrainte de régularité, représentée par le carré de la norme de la dérivée seconde de la fonction de risque de base. Plus spécifiquement, les estimateurs du maximum de vraisemblance pénalisée (MPnLE) pour $\lambda_0(t), \beta$ et θ maximisent (10).

Lorsque la vraisemblance pénalisée est utilisée pour estimer non-paramétriquement des fonctions de régression dans un modèle linéaire, on peut montrer que les estimateurs sont des splines cubiques avec des noeuds placés en

chaque temps d'observation (Silverman, 1985). Cela est vrai aussi dans certaines situations pour un modèle de Cox (Zucker et Karr, 1990). Dans le contexte considéré d'estimation de la fonction de risque, nous n'avons pas cette simplification. Le calcul exact des estimateurs n'est pas possible et le MPnLE pour $\lambda_0(t)$ doit être approché par des splines (ou une autre base de fonctions). Les splines sont des fonctions polynomiales qui sont combinées linéairement de sorte que $\tilde{\lambda}_0(\cdot) = \sum_{i=1}^m \tilde{\eta}_i M_i(\cdot)$, où les $M_i(\cdot)$ sont des M-splines cubiques, i.e. d'ordre 4 (Ramsay, 1988) et $\eta = (\eta_1, \dots, \eta_m)'$ est le vecteur des coefficients des splines. Par intégration directe, on peut aussi obtenir la fonction de risque cumulé : $\tilde{\Lambda}_0(\cdot) = \sum_{i=1}^m \tilde{\eta}_i I_i(\cdot)$ où $I_i(\cdot)$ sont des I-splines définis par $I_i(x) = \int_0^x M_i(u) du$. Il faut remarquer que l'estimateur $\hat{\lambda}(\cdot)$ est le MPnLE, alors que $\tilde{\lambda}(\cdot)$ en est une approximation; il est théoriquement possible de rendre l'erreur d'approximation aussi petite qu'on le souhaite en augmentant le nombre de noeuds.

Le paramètre de lissage κ , va permettre de plus ou moins lisser; il peut être estimé par une méthode approchée de validation croisée en maximisant :

$$\overline{CV(\kappa)} = \frac{1}{n} l_j(\hat{\eta}(\kappa)) - \frac{1}{n} \text{trace} \left(\left[\hat{H}(\hat{\eta}) \right]^{-1} \hat{I}(\hat{\eta}) \right)$$

avec $I(\eta) = E \left(-\frac{\partial^2 l}{\partial \eta^2}(\eta) \right)$ la matrice d'information, $\left(\hat{I}(\hat{\eta}) = -\frac{\partial^2 l}{\partial \eta^2}(\hat{\eta}) \right)$, et $H(\eta) = I(\eta) + 2\kappa\Omega$, et $\hat{H}(\hat{\eta}) = -\frac{\partial^2 pl}{\partial \eta^2}(\hat{\eta}) = \hat{I}(\hat{\eta}) + 2\kappa\Omega$, avec $\Omega = \int \left(\frac{\partial^2 M}{\partial u^2}(u) \right)' \frac{\partial^2 M}{\partial u^2}(u) du$, et $l_j(\cdot)$ la log-vraisemblance pour le sujet j , $\hat{\eta}$ l'estimateur du vecteur des coefficients des splines.

Autrement, en fixant un nombre de degrés de liberté (ddl) nécessaire pour estimer la fonction de risque, on peut utiliser la relation qui associe le nombre de ddl du modèle et le paramètre de lissage pour évaluer κ : $ddl = \text{trace} \left(\left[\hat{I}(\hat{\eta}) + 2\kappa\Omega \right]^{-1} \hat{I}(\hat{\eta}) \right)$. Par exemple, pour estimer une droite, on choisira plutôt un nombre de ddl égal à 2.

La variance des paramètres $\zeta' = (\eta_1, \dots, \eta_m; \beta_1, \dots, \beta_p; \theta)$ peut être estimée directement par $\hat{H}^{-1} = \hat{H}^{-1}(\hat{\zeta})$ ou par $\hat{H}^{-1} \hat{I} \hat{H}^{-1} = \hat{H}^{-1}(\hat{\zeta}) \hat{I}(\hat{\zeta}) \hat{H}^{-1}(\hat{\zeta})$ si on s'inspire des travaux de Gray (1992), avec $I(\zeta) = E \left(-\frac{\partial^2 l}{\partial \zeta^2}(\zeta) \right)$ la matrice d'information $\left(\hat{I}(\hat{\zeta}) = -\frac{\partial^2 l}{\partial \zeta^2}(\hat{\zeta}) \right)$, et $\hat{H}(\hat{\zeta}) = -\frac{\partial^2 pl}{\partial \zeta^2}(\hat{\zeta})$. Nous avons démontré la normalité asymptotique de ces estimateurs du maximum de vraisemblance pénalisée (Rondeau *et al.*, 2002). De plus, en comparant deux méthodes d'estimation des paramètres, celle utilisant l'algorithme EM (Nielsen, 1992) à celle utilisant la vraisemblance pénalisée sur la fonction de risque (Rondeau *et al.*, 2002), on constate que sur des échantillons de petites tailles l'estimation de θ est sensiblement meilleur avec la vraisemblance pénalisée.

3.4. Exemples

Nous présentons ici les données publiées par McGilchrist et Aisbett (1991), et qui ont été utilisées par d'autres auteurs pour illustrer l'utilisation des modèles à fragilités. Chaque observation de survie correspond au délai jusqu'à la survenue d'une infection, après la pause d'un cathéter, chez des patients atteints d'insuffisance rénale. Les cathéters peuvent être retirés pour une raison différente d'une infection; dans ce cas, les données sont censurées. Il y a 38 patients, chacun présentant exactement deux observations. Nous sommes donc en présence de données répétées et les observations pour chaque individu sont potentiellement corrélées. Nous avons comparé la méthode d'estimation par vraisemblance pénalisée sur les variables de fragilité, proposée par Therneau *et al.*(2001) (en utilisant le logiciel R), à celle proposée par Rondeau *et al.*(2002) qui utilise une méthode d'estimation par vraisemblance pénalisée sur la fonction de risque. Nous utilisons un modèle à risques proportionnels suivi d'un modèle à fragilités partagées. Si l'on est en présence d'une corrélation des données pour chaque sujet, on s'attend à obtenir une sous-estimation des paramètres de régression dans un modèle à risques proportionnels par rapport à un modèle à fragilités.

Les résultats (Tableau 1) mettent en évidence une corrélation intra-sujet significative. Ainsi, les valeurs des coefficients sont plus élevées dans les modèles à fragilités par rapport au modèle à risques proportionnels, sous l'hypothèse d'indépendance des temps de survie. Notamment pour l'effet du sexe, significativement associé à la survenue d'une infection, on peut noter une nette sous-estimation de son effet dans un modèle qui ne tient pas compte des corrélations intra-sujets. De plus on peut montrer analytiquement que puisque $I \leq H$, on a $H^{-1}IH^{-1} \leq H^{-1}$. On obtient bien dans l'exemple, comme prévu, des écarts-type plus faibles en utilisant l'estimateur sandwich $\hat{H}^{-1}\hat{I}\hat{H}^{-1}$ plutôt que directement \hat{H}^{-1} .

Dans cet exemple, nous avons également estimé les paramètres en utilisant une loi gaussienne pour les variables de fragilité et une méthode d'estimation par vraisemblance pénalisée sur les variables de fragilité (avec le logiciel R). Les résultats étaient proches de ceux obtenus avec une variable de fragilité gamma, (pour l'âge $\beta = 0.0047$, SE = 0.0125 et pour le sexe : $\beta = -1.41$, SE=0.445). La meilleure vraisemblance était celle obtenue avec la loi gaussienne (-164.13 versus -164.52 pour la loi gamma).

En outre, la vraisemblance pénalisée permet d'estimer $\lambda_0(t)$, mais dans cet exemple, à cause du faible nombre de données, la pénalisation choisie par validation croisée est importante et l'estimation de la fonction de risque est linéaire en fonction du temps.

4. Modèles à fragilités corrélées

Une autre structure de modèle à fragilités a été proposée pour tenir compte à la fois d'une hétérogénéité due à des variables individuelles non observées et d'une corrélation entre certains individus. Ce modèle, dit à fragilités corrélées ou modèle additif à fragilités, a été développé par plusieurs auteurs (Pickles et

MODÉLISATION DE LA FRAGILITÉ EN SURVIE

TABLEAU 1. – Comparaison de différentes méthodes d'estimation dans un modèle à risques proportionnels et dans un modèle à fragilités partagées.

	MODÈLE À RISQUES PROPORTIONNELS			MODÈLE À FRAGILITÉS PARTAGÉES		
	Vraisemblance partielle			Vraisemblance pénalisée sur la variable de fragilité		
	coeff	SE(coeff)*	SE2(coeff)**	coeff	SE(coeff)*	SE2(coeff)**
age	0.0020	0.0092	.	0.0052	0.0119	0.0088
sexe	-0.829	0.299	.	-1.59	0.461	0.352
fragilité	-	-	-	0.412	.	.
	Vraisemblance pénalisée sur la fonction de risque			Vraisemblance pénalisée sur la fonction de risque		
	coeff	SE(coeff)*	SE2(coeff)**	coeff	SE(coeff)*	SE2(coeff)**
age	0.0076	0.0092	0.0092	0.0077	0.0118	0.0116
sexe	-0.706	0.318	0.326	-1.70	0.448	0.388
fragilité	-	-	-	0.447	0.171	0.163

* utilisant l'estimateur $\widehat{H}^{-1}(\hat{\theta})$ de la variance de $\hat{\theta}$,

** utilisant l'estimateur $H^{-1}I\widehat{H}^{-1}(\hat{\theta})$ de la variance de $\hat{\theta}$.

Crouchley, 1994; Yashin *et al.*1995; Petersen, 1998; Korsgaard et Andersen, 1998) qui cherchaient par exemple à distinguer l'effet des facteurs de risque environnementaux par rapport aux facteurs génétiques sur la survie des jumeaux.

La formulation générale du modèle à fragilités corrélées se présente de la manière suivante pour le sujet j :

$$\lambda_j(t|Z^{(j)}) = Z^{(j)}\lambda_0(t) \exp(\beta'X_j) \tag{11}$$

où $\lambda_0(t)$ est la fonction de risque de base en un temps t et $Z^{(j)}$ est une variable aléatoire spécifique au sujet j . Ces variables aléatoires sont corrélées selon une structure de corrélation à définir. Dans le cas particulier où les $Z^{(j)}$ sont indépendants, on se ramène à un modèle de surdispersion.

Yashin *et al.*(1995) ont par exemple proposé d'appliquer un modèle à fragilités corrélées sur une cohorte de jumeaux. Pour cela ils utilisent une décomposition de la variable de fragilité en une somme de deux variables de fragilité, une partagée par plusieurs individus d'un même groupe, l'autre étant non partagée. La structure de corrélation additive est : $Z_i^{(1)} = Z_{i0} + Z_{i1}$ et $Z_i^{(2)} = Z_{i0} + Z_{i2}$. Ces variables aléatoires dans chaque groupe sont corrélées. Par contre les variables aléatoires Z_{i0} , Z_{i1} , et Z_{i2} sont indépendantes et distribuées selon des lois gamma de paramètres (c, θ) , (c^*, θ) et (c^*, θ) respectivement.

Ainsi, les variables aléatoires $Z_i^{(j)}$ suivront une loi gamma de paramètres $(c + c^*, \theta)$. Comme précédemment, pour rendre $\lambda_0(t)$ identifiable, on suppose

$1/\theta = c + c^*$, ainsi l'espérance des fragilités sera égale à 1 (soit $E(Z_i^{(j)}) = (c+c^*)\theta = 1$) et la variance des $Z_i^{(j)}$ sera égale à θ (soit $(c+c^*)\theta^2 = \theta$). Lorsque $\text{var}(Z_{ij}) = 0$ (pour $j \neq 0$), le modèle devient le modèle à fragilités partagées. La variable Z_{i0} induit une corrélation des individus dans chaque groupe et peut refléter l'ensemble des facteurs de risque génétiques ou environnementaux communs aux individus d'un même groupe. Les variables Z_{ij} (pour $j \neq 0$) traduisent une éventuelle hétérogénéité entre individus même après avoir pris en compte des facteurs de risque génétiques ou environnementaux communs. Cela peut correspondre par exemple à un environnement non partagé (ex : habitudes alimentaires).

Le modèle à fragilités corrélées conduit à une expression de la vraisemblance plus compliquée que celle utilisée dans le modèle à fragilités partagées (Petersen, 1998). En supposant que la censure est indépendante et non informative pour Z , la vraisemblance sur les données non observées prend la forme :

$$\prod_{i=1}^G \left\{ \prod_{j=1}^{n_i} \left[\lambda_{ij}(Y_{ij}|Z_i^{(j)})^{\delta_{ij}} \exp \left(-Z_i^{(j)} \Lambda_{ij}(Y_{ij}) \right) p(Z_{ij}; c^*, \theta) \right] p(Z_{i0}; c, \theta) \right\}$$

où $p(\cdot; c, \theta)$ est la densité de probabilité d'une loi gamma de paramètres (c, θ) .

L'algorithme EM peut être également utilisé pour estimer les paramètres selon la même procédure décrite pour les modèles à fragilités partagées. Il faudra cependant utiliser des calculs récursifs pour évaluer des espérances conditionnelles des variables de fragilité (utilisée dans l'étape E).

Murphy (1994; 1995) a démontré la consistance et la normalité asymptotique des estimateurs non-paramétriques du maximum de vraisemblance dans le modèle à fragilités partagées sans variables explicatives, c'est-à-dire lorsque les paramètres à estimer sont la fonction de risque cumulé de base et la variance des variables de fragilité (de loi gamma). Plus récemment, Parner (1998) a étendu ces résultats au modèle à fragilités corrélées avec des variables explicatives.

Les modèles à fragilités trouvent de nombreuses applications dans le domaine de la génétique. Les variables aléatoires peuvent par exemple traduire un facteur génétique non observé qui affecte un groupe d'animaux. Ducrocq et Casella (1996) proposent d'utiliser une approche bayésienne pour estimer les paramètres de la distribution des effets aléatoires dans des modèles à fragilités corrélées de la forme :

$$\lambda_{ij}(t|\eta, \rho) = \lambda_0(t) \exp(\mathbf{W}'_{ij}\eta) \tag{12}$$

Ils supposent une distribution de Weibull pour la fonction de risque de base $\lambda_0(t) = \lambda\rho(\lambda t)^{\rho-1}$, avec $\eta = \begin{pmatrix} \beta \\ Z \end{pmatrix}$ le vecteur des effets fixes et des effets aléatoires, et $\mathbf{W}'_{ij} = (X'_{ij}, S'_{ij})$ la matrice de « design » connue, comme dans un modèle linéaire mixte. Lorsque l'on se place dans le cas d'un modèle avec un seul effet aléatoire spécifique à chaque groupe $v_i = e^{Z_i}$, avec une loi a priori

gamma de paramètre de variance θ , alors Z_i a une loi a priori log-gamma de densité $g(z) = \frac{\exp((z - e^z)/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}}$. La log-vraisemblance marginale s'obtient par intégration analytique simple de la vraisemblance conjointe comme dans la formule (5).

Ils considèrent des cas plus complexes, avec par exemple des variables explicatives dépendantes du temps, ou des effets aléatoires corrélés à travers une matrice de corrélation connue. Ils proposent d'utiliser une marginalisation basée sur une intégration de Laplace, une technique liée aux approximations points-selles.

5. Conclusion

Les modèles à fragilités sont intéressants pour modéliser l'hétérogénéité de la population qui ne peut pas être expliquée par des facteurs connus et mesurables. Cette hétérogénéité peut être présente même si les observations sont indépendantes; cependant, c'est dans le cas où l'on a des données corrélées que l'on peut valablement estimer des paramètres mesurant cette hétérogénéité : dans les modèles à fragilités partagées, la variance de la variable de fragilité est un indice direct de l'hétérogénéité. Les modèles à fragilités corrélées sont plus souples que les modèles à fragilités partagées et permettent d'aborder des structures de corrélation plus complexes que l'on trouve en particulier en génétique. Notons au passage que la génétique semble être le domaine d'application le plus prometteur des modèles à fragilités ainsi que le montrent les applications de Petersen (1998) et de Ducrocq et Casella (1996) pour des modèles de corrélations de traits ainsi que des applications possibles en liaison génétique (Commenges, 1995). En général il semble que l'hétérogénéité inexplicée d'origine géographique soit plus faible et souvent non détectable (Rondeau *et al.*, 2002). Des problèmes numériques restent à surmonter pour appliquer plus largement ces modèles. Si les modèles à fragilités gamma ou log-normale partagées semblent applicables aujourd'hui, les modèles utilisant d'autres distributions et *a fortiori* les modèles faisant intervenir plusieurs variables de fragilité posent encore des problèmes.

Des termes de fragilité ont été également utilisés dans les modèles de vie accélérée, pour étudier par exemple l'hétérogénéité liée à des variables explicatives omises (Keiding *et al.*, 1997).

Une extension des modèles à fragilités est de considérer que la variable de fragilité dépend du temps. Cela revient à considérer un processus latent. Des modèles à processus latent sont développés, en particulier pour la modélisation conjointe de marqueurs et d'événements (Henderson *et al.*, 2001). Cependant ces modèles sont a priori plus utiles pour aborder des problèmes multivariés que des problèmes à événements répétés.

RÉFÉRENCES

- AALEN O.O. (1988), Heterogeneity in survival analysis, *Statistics in Medicine*, **7**, 1121-1137.
- AALEN O.O. (1994), Effects of frailty in survival analysis, *Statistical methods in medical research*, **3**, 227-243.
- CLAYTON D.G. (1978), A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**, 141-151.
- CLAYTON D.G., CUZICK J. (1985), Multivariate generalizations of the proportional hazards model (with discussion), *Journal of the Royal Statistical Society A*, **148**, 82-117.
- CLAYTON D.G. (1991), A monte carlo method for bayesian inference in frailty models, *Biometrics*, **47**, 467-485.
- COMMENGES D. (1995), Developments of the WPC approach and application to linkage analysis of breast cancer, *Genetic Epidemiology*, **12**, 853-857.
- COMMENGES D., ANDERSEN P.K. (1995), Score test of homogeneity for survival data, *Lifetime Data Analysis*, **1**, 145-156.
- DUCROCQ V., CASELLA G. (1996), A Bayesian analysis of mixed survival models, *Genetics selection evolution*, **28**, 505-529.
- GRAY R.J. (1992), Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis, *Journal of the American Statistical Association*, **87**, 942-951.
- HA I.D., LEE Y., SONG J.K. (2001), Hierarchical likelihood approach for frailty models, *Biometrika*, **88**, 233-243.
- HARVILLE D. (1977), Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association*, **72**, 320-340.
- HENDERSON R., DIGGLE P., DOBSON A. (2001), Joint modelling of longitudinal measurements and event time data, *Biostatistics*, **01**, 465-480.
- HOUGAARD P. (1986), A class of multivariate failure time distributions, *Biometrika*, **73**, 671-678.
- HOUGAARD P. (1995), Frailty models for survival data, *Lifetime Data Analysis*, 255-273.
- HOUGAARD P. (2000), Analysis of Multivariate Survival Data, *Springer-Verlag New-York*.
- JOHANSEN S. (1983), An extension of Cox's regression model, *International Statistical Review*, **51**, 165-174.
- JOLY P., COMMENGES D., LETENNEUR L. (1998), A penalized likelihood approach for arbitrarily censored and truncated data : application to age-specific incidence of dementia, *Biometrics*, **54**, 185-194.
- KEIDING N., ANDERSEN P.K., KLEIN J.P. (1997), The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates, *Statistics in Medicine*, **16**, 215-224.
- KLEIN J.P. (1992), Semiparametric estimation of random effects using the Cox model based on the EM algorithm, *Biometrics*, **48**, 795-806.
- KORSGAARD I.R., ANDERSEN A.H. (1998), The additive genetic gamma frailty model, *Scandinavian Journal of Statistics*, **25**, 255-269.

- LEE Y., NELDER J.A. (1996), Hierarchical generalized linear models, *Journal of the Royal Statistical Society B*, 58, 619-678.
- MCGILCHRIST C.A., AISBETT C.W. (1991), Regression with frailty in survival analysis, *Biometrics*, 47, 461-466.
- MURPHY S.A. (1994), Consistency in a proportional hazards model incorporating a random effect, *The Annals of Statistics*, 22, 712-731.
- MURPHY S.A. (1995), Asymptotic theory for the frailty model, *The Annals of Statistics*, 23, 182-198.
- NIELSEN G.G., GILL R.D., ANDERSEN P.K., SORENSEN T.I.A. (1992), A counting process approach to maximum likelihood estimation in frailty models, *Scandinavian Journal of Statistics*, 19, 25-43.
- O'SULLIVAN F. (1988), Fast computation of fully automated log-density and log-hazard estimators, *SIAM Journal on Scientific Computing*, 9, 363-379.
- PARNER E. (1997), Inference in semiparametric frailty models. Ph.D. dissertation, University of Aarhus, Denmark.
- PARNER E. (1998), Asymptotic theory for the correlated gamma-frailty model, *The Annals of Statistics*, 26, 183-214.
- PETERSEN J.H. (1998), An additive frailty model for correlated life times, *Biometrics*, 54, 646-661.
- PICKLES A., CROUCHLEY R. (1994), Generalizations and applications of frailty models for survival and event data, *Statistical methods in medical research*, 3, 263-278.
- RAMSAY J.O. (1988), Monotone regression splines in action, *Statistical Science*, 3, 425-461.
- RIPATTI S., PALMGREN J. (2000), Estimation of multivariate frailty models using penalized partial likelihood, *Biometrics*, 56, 1016-1022.
- RONDEAU V., COMMENGES D., JOLY P. (2002), Maximum penalized likelihood estimation in frailty models, *In press, Lifetime Data Analysis*.
- SILVERMAN B.W. (1985), Some aspects of the spline smoothing approach to non-parametric regression curve fitting, *Journal of the Royal Statistical Society B*, 47, 1-52.
- THERNEAU T.M., GRAMBSCH P.M. (2001), Modeling Survival Data : Extending the Cox Model, *Springer-Verlag New-York*.
- TIETZE C. (1968), Fertility after discontinuation of intrauterine and oral contraception, *International Journal of Fertility*, 13, 385-389.
- WEI L.J., LIN D.Y., WEISSFELD L. (1989), Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association*, 84, 1065-1073.
- YASHIN A.I., VAUPEL J.W., IACHINE I.A. (1995), Correlated individual frailty : an advantageous approach to survival analysis of bivariate data, *Mathematical Population studies*, 5, 145-159.
- ZUCKER D.M., KARR A.F. (1990), Nonparametric survival analysis with time-dependent covariate effects : A penalized partial likelihood approach, *The Annals of Statistics*, 18, 329-353.